

CLAIMS

What is claimed is:

- 1 1. A computerized method of storing table data comprising:
2 parsing the table data into columns of values;
3 formatting each column into a data stream; and
4 directing a storage device to store each data stream as a continuous strip of
5 compressed data that extends across page boundaries.
- 1 2. The computerized method of claim 1 further comprising:
2 partitioning each column into groups of values based on a primary key for the table
3 data; and
4 formatting each group of values into a data stream.
- 1 3. The computerized method of claim 1, wherein formatting each column comprises:
2 compressing the values in the column.
- 1 4. The computerized method of claim 3, wherein compressing the values comprises:
2 creating a code for each value in the column; and
3 replacing each value with the corresponding code.
- 1 5. The computerized method of claim 4, wherein creating a code for each value
2 comprises:
3 creating a plurality of entries, one entry for each value in the column; and
4 deriving the code from a location for the corresponding entry within the plurality of
5 entries.
- 1 6. The computerized method of claim 4, wherein creating a code for each value

2 comprises:

3 determining a number of occurrences of each value in the column; and

4 deriving the code for each value from the corresponding number of occurrences.

1 7. The computerized method of claim 4, wherein creating a code for each value

2 comprises:

3 creating a plurality of entries, one entry for each value in the column;

4 storing a number of occurrences of each value in the column in the corresponding

5 entry; and

6 deriving the code for each value from the corresponding number of occurrences.

1 8. The computerized method of claim 7 further comprising:

2 directing the storage device to store the plurality of entries in conjunction with the

3 corresponding continuous strip of data.

1 9. The computerized method of claim 7 further comprising:

2 directing the storage device to store the plurality of entries in a header for the

3 corresponding continuous strip of data.

1 10. The computerized method of claim 4, wherein compressing the values further

2 comprises:

3 encoding the codes in the column according to an encoding table.

1 11. The computerized method of claim 1, wherein formatting each column into a data

2 stream comprises:

3 formatting multiple columns into a single data stream.

1 12. The computerized method of claim 11, wherein formatting multiple columns

2 comprises linearly concatenating a series of rows, each row comprising one value from
3 each of the multiple columns.

1 13. The computerized method of claim 11, wherein formatting multiple columns
2 comprises linearly concatenating the multiple columns.

1 14. A computer-readable medium having executable instructions to cause a computer
2 to execute a method comprising:
3 parsing table data into columns of values;
4 formatting each column into a data stream; and
5 transferring each data stream to a storage device for storage as a continuous strip of
6 compressed data that extends across page boundaries.

1 15. The computer-readable medium of claim 14, wherein the method further
2 comprises:
3 partitioning each column into groups of values based on a primary key for the table
4 data; and
5 formatting each group of values into a data stream.

1 16. The computer-readable medium of claim 14, wherein the method further comprises
2 compressing the values in a column when formatting the column.

1 17. The computer-readable medium of claim 16, wherein the method further
2 comprises:
3 creating a code for each value in the column; and
4 replacing each value with the corresponding code when compressing the values in
5 the column.

1 18. The computer-readable medium of claim 17, wherein the method further
2 comprises:
3 creating a plurality of entries, one entry for each value in the column; and
4 deriving the code from a location for the corresponding entry within the plurality of
5 entries.

1 19. The computer-readable medium of claim 17, wherein the method further
2 comprises:
3 determining a number of occurrences of each value in the column; and
4 deriving the code for each value from the corresponding number of occurrences.

1 20. The computer-readable medium of claim 17, wherein the method further
2 comprises:
3 creating a plurality of entries, one entry for each value in the column;
4 storing a number of occurrences of each value in the column in the corresponding
5 entry; and
6 deriving the code for each value from the corresponding number of occurrences.

1 21. The computer-readable medium of claim 20, wherein the method further
2 comprises:
3 directing the storage device to store the plurality of entries in conjunction with the
4 corresponding continuous strip of data.

1 22. The computer-readable medium of claim 20, wherein the method further
2 comprises:
3 directing the storage device to store the plurality of entries in a header for the
4 corresponding continuous strip of data.

1 23. The computer-readable medium of claim 17, wherein the method further
2 comprises:
3 encoding the codes in the column according to an encoding table.

1 24. The computer-readable medium of claim 14, wherein the method further
2 comprises:
3 formatting multiple columns into a single data stream.

1 25. The computer-readable medium of claim 24, wherein the method further comprises
2 linearly concatenating a series of rows, each row comprising one value from each of the
3 multiple columns, when formatting the multiple columns.

1 26. The computer-readable medium of claim 24, wherein the method further comprises
2 linearly concatenating the multiple columns when formatting the multiple columns.

1 27. A computer system comprising:
2 a processing unit;
3 a memory coupled to the processing unit through a bus;
4 a storage device coupled to the processing unit through a bus;
5 a data storing process executed from the memory by the processing unit to cause
6 the processing unit to parse table data into columns of values, to format each column into a
7 data stream, and to direct the storage device to store the data stream as a continuous strip
8 of compressed data that extends across page boundaries.

1 28. The computer system of claim 27, wherein the data storing process further causes
2 the processing unit to partition each column into groups of values based on a primary key
3 for the table data and to format each group of values into a data stream.

1 29. The computer system of claim 27, wherein the data storing process further causes
2 the processing unit to compress the values in a column when formatting the column.

1 30. The computer system of claim 29, wherein the data storing process further causes
2 the processing unit to create a code for each value in the column and to replace each value
3 with the corresponding code when compressing the values in the column.

1 31. The computer system of claim 30, wherein the data storing process further causes
2 the processing unit to create a plurality of entries, one entry for each value in the column
3 and to derive the code from a location for the corresponding entry within the plurality of
4 entries.

1 32. The computer system of claim 30, wherein the data storing process further causes
2 the processing unit to determine a number of occurrences of each value in the column and
3 to derive the code for each value from the corresponding number of occurrences.

1 33. The computer system of claim 30, wherein the data storing process further causes
2 the processing unit to create a plurality of entries, one entry for each value in the column,
3 to store a number of occurrences of each value in the column in the corresponding entry,
4 and to derive the code for each value from the corresponding number of occurrences.

1 34. The computer system of claim 33, wherein the data storing process further causes
2 the processing unit to direct the storage device to store the plurality of entries in
3 conjunction with the corresponding continuous strip of data.

1 35. The computer system of claim 33, wherein the data storing process further causes
2 the processing unit to direct the storage device to store the plurality of entries in a header
3 for the corresponding continuous strip of data.

1 36. The computer system of claim 30, wherein the data storing process further causes
2 the processing unit to encode the codes in the column according to an encoding table when
3 compressing the values in the column.

1 37. The computer system of claim 27, wherein the data storing process further causes
2 the processing unit to format multiple columns into a single data stream.

1 38. The computer system of claim 37, wherein the data storing process further causes
2 the processing unit to linearly concatenate a series of rows, each row comprising one value
3 from each of the multiple columns, when formatting the multiple columns.

1 39. The computer system of claim 37, wherein the data storing process further causes
2 the processing unit to linearly concatenate the multiple columns when formatting the
3 multiple columns.

1 40. A data storing system comprising:
2 a plurality of compute nodes coupled to a data source, to receive table data from
3 the data source and to parse the table data into columns of values; and
4 a plurality of storage nodes, each storage node comprising a storage device and
5 coupled to the plurality of compute nodes to receive the columns of values from the
6 compute nodes, to format each column into a data stream, and to direct the storage device
7 to store the data stream as a continuous strip of compressed data that extends across page
8 boundaries.

1 41. The data storing system of claim 40, wherein the plurality of compute nodes are
2 further operable to partition each column into groups of values based on a primary key for
3 the table data, and each storage node is further operable to format a group of values into a
4 data stream.

1 42. The data storing system of claim 40, wherein each storage node is further operable
2 to compress the values in a column when formatting the column.

1 43. The data storing system of claim 42, wherein each storage node is further operable
2 create a code for each value in a column and to replace each value with the corresponding
3 code when compressing the values in the column.

1 44. The data storing system of claim 42, wherein each storage node is further operable
2 to create a plurality of entries, one entry for each value in the column, to store a number of
3 occurrences of each value in the column in the corresponding entry, and to derive the code
4 for each value from the corresponding number of occurrences.

1 45. The data storing system of claim 44, wherein each storage node is further operable
2 to direct the storage device to store the plurality of entries in conjunction with the
3 corresponding continuous strip of data.

1 46. The data storing system of claim 44, wherein each storage node is further operable
2 to direct the storage device to store the plurality of entries in a header for the
3 corresponding continuous strip of data.

1 47. The data storing system of claim 43, wherein each storage node is further operable
2 to encode the codes in the column according to an encoding table when compressing the
3 values in the column.

1 48. The data storing system of claim 40, wherein each storage node is further operable
2 to format multiple columns into a single data stream.

1 49. The data storing system of claim 48, wherein each storage node is further operable

2 to linearly concatenate a series of rows, each row comprising one value from each of the
3 multiple columns, when formatting the multiple columns.

1 50. The data storing system of claim 48, wherein each storage node is further operable
2 to linearly concatenate the multiple columns when formatting the multiple columns.

1 51. The data storing system of claim 40, wherein one of the plurality of compute nodes
2 acts as a master to receive the table data from the data source and to transfer the table data
3 and instructions for storing the table data to the other compute nodes.

1 52. A data structure comprising:
2 a header field containing data representing an identifier for a column of values
3 from a table; and
4 a plurality of data fields containing data representing the values in the column
5 identified by the header field, the plurality of data fields forming a continuous stream of
6 compressed data for storing across page boundaries.

1 53. The data structure of claim 52, further comprising:
2 a plurality of dictionary entries containing data representing each value in the
3 column and data representing a count of the occurrences of the corresponding value in the
4 column identified by the header field, wherein the data in the plurality of data fields are
5 codes derived from the counts of the occurrences of the corresponding values.

1 54. The data structure of claim 53, wherein the header field further contains data
2 representing the plurality of dictionary entries.